Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-03-15 02:23:03
REFERENCE NO: 182

Author Names & Affiliations

- Francesco Paesani - University of California, San Diego

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Theoretical and Computational Chemistry

Title of Submission

Enabling Chemical Accuracy in Molecular-Level Computer Simulations

Abstract (maximum ~200 words).

Molecular-level computer simulations have become indispensable in many research areas, including chemistry, physics, materials science, and biochemistry, and often provide fundamental insights that are otherwise difficult to obtain. Nowadays, computer simulations are used to complement, guide, and sometimes replace experimental measurements, reducing the amount of time and money spent on research to bring ideas from the lab to practical applications.

However, both the realism and the predictive power of a molecular-level computer simulation directly depend on the accuracy with which the interactions between molecules are described. To address the limitations of existing simulation approaches, we have recently developed a new theoretical/ computational methodology, so-called many-body molecular dynamics (MB-MD), which exhibits high accuracy when applied to a variety of molecular systems. Our research focuses on the development and implementation of unique software elements that will enable computer simulations on both CPU and GPU architectures using MB-MD methodology. These software elements will be made publicly available to the scientific community through an integrated platform.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

As attested by the 2013 Nobel Prize in Chemistry awarded to Martin Karplus, Michael Levitt, and Arieh Warshel, computer simulations have become a powerful tool in many research areas, including chemistry, physics, materials science, and biochemistry, and often provide fundamental insights into complex phenomena that are otherwise difficult to obtain. However, both the realism and the predicting power of a computer simulation largely depend on the accuracy with which the molecular interactions and the overall system dynamics are described. While, in principle, correlated wavefunction theory (WFT) can provide the correct description (within chemical accuracy) of molecular

interactions from isolated molecules to condensed-phase systems, the associated computational cost effectively limits the application of correlated WFT methods to single-point calculations on relatively small systems. In this context, it should be mentioned that recent advances in quantum Monte Carlo (QMC) techniques, which have comparable accuracy to correlated WFT methods, have also enabled benchmark single-point calculations for bulk systems.

Despite much recent progress in both WFT and QMC methods, density functional theory (DFT) effectively remains the only viable ab initio approach applicable to the study of condensed-phase systems. However, besides being still computationally expensive, common DFT methods have not yet reached chemical accuracy, particularly for weakly-interacting and hydrogen-bonded systems. As a consequence, current simulation methodologies still lack the necessary accuracy, so-called chemical accuracy, to become reliable predictive tools.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

Data-driven approaches have recently emerged as a promising alternatives to both standard classical computer simulations carried out with force fields based on empirical parameterizations and ab initio methods. Our MB-MD methodology, like other machine-learning-based methods, lies within the family of data-driven approaches. Taking water as a prototypical system that, despite having been studied through computer simulations for more than 50 years, remains a topic of intense debate, our MB-MD methodology overcome existing limitations in classical and ab initio molecular simulations, accurately predicting structural, thermodynamic, dynamical, and spectroscopic properties of water across different phases. More recently, we have extended our MB-MD methodology to the description of ions in aqueous systems. Our initial studies clearly show that DFT models commonly used in simulations of ionic systems are unable to correctly describe charge-dipole interactions between the halide ions and water molecules, mainly because of deficiencies at short range where charge transfer and exchange-repulsion play major roles. On the other hand, by construction, classical (polarizable) force fields underestimate the degree of charge transfer and, therefore, are unable to correctly describe ion-water interactions at short range due to the more covalent nature of these interactions. In contrast, our MB-MD methodology correctly capture the underlying physics of ion-water interactions, independently of the nature of the ion and the system size. Our current theoretical efforts within the MB-MD methodology focus on recasting electronic structure embedding schemes in terms of the many-body formalism as well as on integrating MB-MD in adaptive quantum mechanics / molecular mechanics (adQM/MM) schemes. These new developments will extend the range of applicability of MB-MD to 1) molecules of arbitrary size and 2) reactive systems, which will result in a unified theoretical/computational methodology for predictive simulations of complex phenomena at the molecular level.

The highest barrier that we see along the path that can lead to the full development and implementation of our MB-MD methodology is the access to students and postdocs trained at the intersection of chemistry, physics, and computer science. The last component, in particular, is largely neglected in the current undergraduate and graduate science curricula, and is primarily left to initiatives of individual research groups and/or specialized summer schools and workshops.

The second challenge that we see is the possibility to have access to professional software engineers that could transform software developed by individual research groups into "commercial-quality" software that can efficiently take advantage of modern hardware architectures.

Consent Statement

- “I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode).”